

# Estudio de la Estructura Tridimensional de Proteínas Mediante la Geometría Fractal

Sergio Gómez González<sup>1</sup>, Joan Grau Barceló<sup>2</sup>,  
Ricardo Torres Camara<sup>2</sup>, Francesc Sepulcre Sánchez<sup>3</sup>

<sup>1</sup> Departamento de Expresión Gráfica en la Ingeniería

<sup>2</sup> Unidad Estructural de Mecánica de la Escuela Universitaria de  
Ingenieros Técnicos Industriales de Barcelona

<sup>3</sup> Unidad Estructural de Química Industrial de la Escuela Universitaria de  
Ingenieros Técnicos Industriales de Barcelona

Universidad Politécnica de Catalunya  
Comte d'Urgell 187  
08036 Barcelona - Spain  
e-mail: Francesc.Sepulcre@upc.es

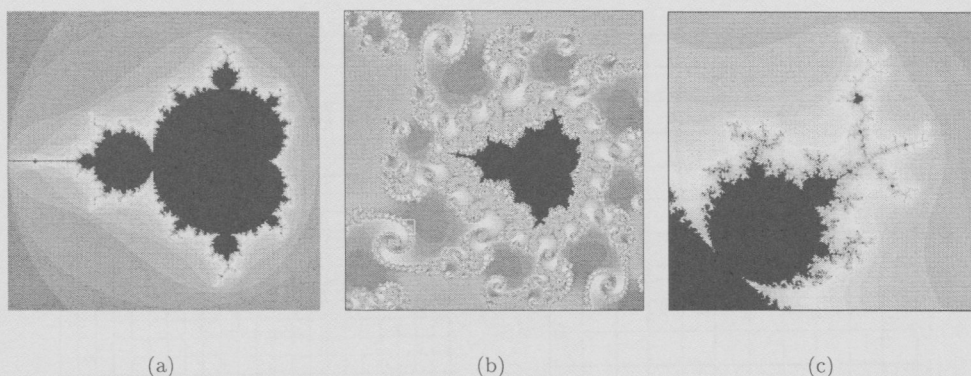
## Resumen

En el presente trabajo se realiza un estudio de caracterización de estructuras secundarias de las proteínas mediante la geometría fractal. Esta geometría se presenta como un método capaz de diferenciar los grados de compactación (estructura terciaria) de las proteínas y sus elementos de estructura secundaria. Las ventajas más importantes de la aplicación de esta geometría son su fácil implementación y la posibilidad de cuantificar, y por tanto diferenciar, el grado de empaquetamiento de estructuras irregulares tridimensionales, como por ejemplo las proteínas.

**Palabras Clave:** Estructura de proteínas, geometría fractal, dimensión fractal, box-counting.

## 1. Introducción

Un fractal viene a ser el producto final que se origina a través de la iteración infinita de un proceso geométrico bien especificado. Este proceso geométrico, que es generalmente de naturaleza muy simple, determina perfectamente la estructura final que, muy frecuentemente, debido a la repetición infinita que se ha efectuado, tiene una complicación aparentemente extraordinaria.



**Figura 1:** Conjunto de Mandelbrot (1980). Obtenido mediante el programa Fractint [2].

Los fractales, por lo tanto, son conjuntos geométricos frecuentemente complicados en apariencia, pero en realidad resultan ser tales que para su descripción, construcción y exploración se requiere muy poca información. También es cierto que para la introducción de los objetos clásicos de la geometría, la circunferencia, el triángulo, ... la información requerida es muy simple, pero ésta proporciona el objeto al que se refiere todo de una vez. Lo característico de las estructuras fractales consiste en la iteración infinita del proceso simple que da lugar al objeto en cuestión.

La esencia de la teoría es que muchas estructuras naturales, como por ejemplo las nubes, que aparentan tener una complejidad extraordinaria, poseen en realidad una misma regularidad geométrica: la denominada *invarianza bajo escala*. Esto significa que, si se analizan estas estructuras a distintas escalas, se encuentra una y otra vez los mismos elementos básicos. Su exploración a distintas escalas encuentra una descripción matemática apropiada mediante el concepto de dimensión fractal.

Todo esto se pone de manifiesto en un modelo matemático descubierto por Mandelbrot en 1980 y conocido hoy como el *Conjunto de Mandelbrot* [6]. Este objeto de gran complejidad presenta una riqueza extraordinaria de formas y estructuras que consta de fragmentos geométricos de orientación y tamaño variable, pero de aspecto similar (ver Figura 1).

## 2. Dimensión fractal

Una de las características principales de cualquier objeto fractal es su dimensión fractal ( $d_f$ ), que mide su grado de irregularidad e interrupción. No obstante, al contrario de la dimensión Euclídea, la dimensión fractal puede muy bien ser una fracción simple, como  $1/2$  ó  $5/3$ , e incluso un número irracional, como  $\log 4 / \log 3$ ,  $-1.2658$  ó  $\pi$ . Así resulta útil decir que para ciertas curvas planas muy irregulares la dimensión fractal está entre 1 y 2, o decir que para superficies muy hojaldradas e irregulares la dimensión fractal es intermedia entre 2 y 3.

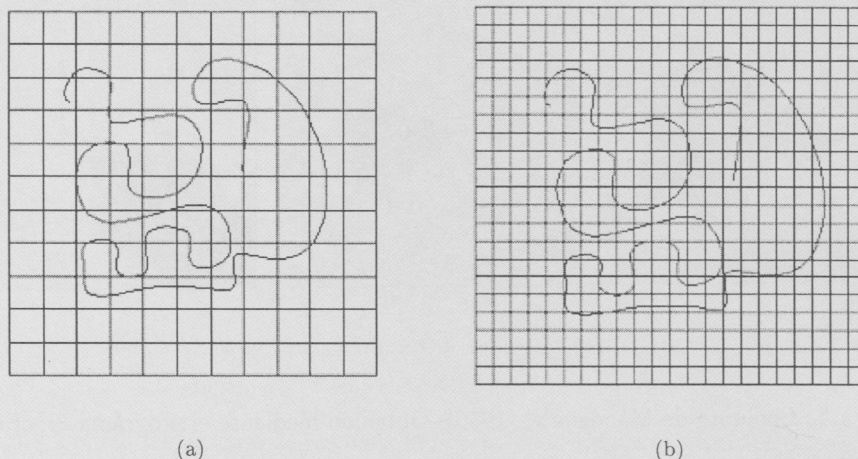


Figura 2: Ejemplo de una curva no rectificable.

Uno de los métodos más extendidos en el cálculo de la dimensión fractal es el *box-counting* o recuento por cajas. La idea central por la que se llega a este concepto de dimensión consiste en diseñar un proceso de medida a una cierta escala  $\delta$ , que ignora las irregularidades de tamaño menor y estudia cómo varía dicha medida a medida que  $\delta$  tiende a cero.

Para ilustrar mejor el concepto de la dimensión fractal obtenida por el método de recuento por cajas podemos imaginarnos una curva  $E$ , no necesariamente rectificable. Si reticulamos el plano con cuadrados de lado  $\delta$  (ver Figura 2) podemos definir " $\delta$  longitud de  $C$ " (longitud que aprecia solamente detalles de tamaño menor que  $\delta$ ) como la suma de las longitudes de los lados de los cuadrados del mallado que cortan a la curva  $C$ , y así, si el número de éstos es  $N_\delta(C)$ , tal  $\delta$ -longitud sería:

$$L_\delta(C) = N_\delta(C)\delta \quad (1)$$

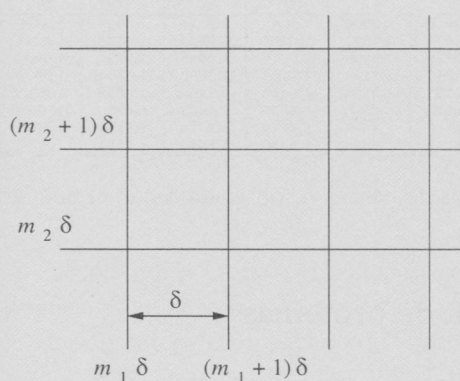
Si la curva es rectificable, cabe esperar que cuando  $\delta$  tiende a cero el número  $L_\delta(C)$  tienda a la longitud de dicha curva, pero se si trata de una curva que mirada a cualquier escala sigue mostrando irregularidades, es decir una curva de dimensión mayor que uno, cuando hagamos tender  $\delta$  a cero  $L_\delta(C)$  tenderá a infinito. La velocidad con la que lo hace, que de alguna forma da el grado de irregularidad, o más bien el tamaño de la curva, puede ser medido por el número  $s$  para el cual  $L_\delta(C)$  y  $\delta^{-s}$  crecen con velocidades comparables, simbólicamente cuando  $\delta \rightarrow 0$ .

$$L_\delta(C) \sim K\delta^{-s} \quad (2)$$

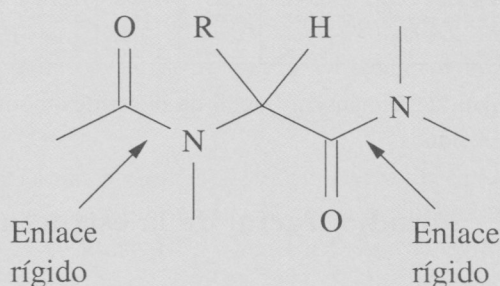
con  $K$  y  $s$  constantes. Por lo tanto, tomando logaritmos,  $\log L_\delta(C) \approx \log \delta$  para  $\delta$  pequeños. O en términos más precisos

$$S = \lim_{\delta \rightarrow 0} \left( \frac{\log L_\delta(C) - \log K}{-\log \delta} \right) = \lim_{\delta \rightarrow 0} \left( \frac{\log L_\delta(C)}{-\log \delta} \right) \quad (3)$$





(a) Ejemplo de reticulado bidimensional.



(b) Enlace peptídico entre dos aminoácidos. El carbono del grupo metileno sustituido, es decir, el que está enlazado al grupo R, es el llamado  $C_{\alpha}$ .

**Figura 3:** Ejemplos de reticulado bidimensional y enlace peptídico entre dos aminoácidos.

ya que el numerador y el denominador tienden a infinito y  $\log K$  es constante. Esto motiva la introducción de un nuevo concepto de dimensión.

En  $\mathbb{R}^n$  consideramos la  $\delta$ -malla formada por la colección de "cubos" de la forma:

$$[m_1\delta, (m_1 + 1)\delta] \times \dots \times [m_n\delta, (m_n + 1)\delta] \quad (4)$$

donde  $m_1 \dots m_n$  son enteros.

Si estamos por ejemplo en  $\mathbb{R}^2$ , se tiene la malla que se ilustra en la Figura 3(a).

Se define la dimensión del recuento por cajas de un conjunto  $E$  como

$$\dim_B E = \lim \left( \frac{\log N_{\delta}(E)}{-\log \delta} \right) \quad (5)$$

donde  $N_{\delta}(E)$  es el número de cubos de la  $\delta$ -malla que corta a  $E$ , en el caso de que dicho límite exista.

En los fenómenos físicos o en las estructuras físicas reales que se pueden describir mediante un modelo fractal (agregados moleculares, curvas de separación de medios, difusión, polímeros, etc.) no tiene sentido la expresión  $\delta \rightarrow 0$  y lo que se hace en la práctica es calcular  $N_{\delta}(E)$  para valores pequeños de  $\delta$ , siendo el término "pequeño" variable según el contexto.

En tales casos, si se representa en un sistema de ejes coordenados el número  $\log N_{\delta}(E)$  en función de  $-\log \delta$ , resulta, cuando  $\delta$  varía dentro de valores adecuados a cada caso, un conjunto de puntos que se aproximan a una recta cuya pendiente nos da la dimensión fractal del recuento por cajas.



Aminoácido N° de residuos	Dimensión Fractal ( $d_f$ ). Conformaciones $\alpha$ .									
	10	20	30	40	50	60	70	80	90	100
Alanina (Ala)	1,2063	1,2122	1,3247	1,1946	1,2660	1,3036	1,3241	1,3454	1,3690	1,3786
Leucina (Leu)	0,7086	1,3225	1,4780	1,5484	1,5929	1,6079	1,6217	1,6295	1,6345	1,6224
Fenilalanina (Phe)	0,9998	1,3311	1,5591	1,6157	1,6738	1,6868	1,6918	1,6947	1,7129	1,6641
Tirosina (Tyr)	1,4153	1,4834	1,5646	1,6460	1,7010	1,7154	1,7107	1,7083	1,5894	1,5939
Triptófano (Trp)	1,4850	1,4127	1,6339	1,6878	1,5285	1,3583	1,5912	1,1461	1,6390	1,6253
Cisteína (Cys)	0,9092	1,2313	1,3840	1,2429	1,2970	1,3370	1,3835	1,4022	1,4259	1,4383
Metionina (Met)	0,9754	1,3352	1,4530	1,3008	1,3543	1,3695	1,4478	1,4746	1,5133	1,5972
Histidina (His)	0,9904	1,3998	1,5243	1,5863	1,4293	1,4690	1,5020	1,5242	1,5581	1,5590

Tabla 1: Dimensión fractal de diferentes modelos de hélices  $\alpha$ , obtenida según el número de residuos.

### 3. Estudio fractal de la estructura de proteínas

#### 3.1. Estructura secundaria de las proteínas

La estructura secundaria de una proteína se define como la conformación local de su cadena polipeptídica (*backbone*). Aproximadamente una tercera parte de los enlaces polipeptídicos de una proteína son rígidos, por lo que dicha cadena puede imaginarse como una serie de planos rígidos separados por grupos metileno sustituidos (Figura 3(b)).

La rigidez de estos enlaces impone ciertas restricciones al número de conformaciones espaciales que la cadena polipeptídica puede adoptar. La conformación de una cadena polipeptídica queda, por tanto, definida en función de los ángulos que adopten los enlaces  $N - C_\alpha$   $\Phi(\Phi)$  y  $C_\alpha - C$   $\Psi(\Psi)$ , que no son rígidos.

Los valores que pueden tomar estos ángulos, que están restringidos por las interacciones electrostáticas y los impedimentos estéricos que aparecen en la proteína, se suman en los diagramas de Ramachandran (Figura 6). Así, en estos diagramas se diferencian zonas que corresponden a los valores de  $\Phi$  y  $\Psi$  permitidos, y otras zonas que corresponden a los valores prohibidos de estos ángulos.

En función de los valores de estos ángulos se pueden diferenciar principalmente dos tipos de estructuras secundarias en las proteínas (ver Figura 6): hélices  $\alpha$  ( $\Phi = -57^\circ$ ;  $\Psi = -47^\circ$ , y 3.6 aminoácidos por vuelta de hélice) y láminas  $\beta$  ( $\Phi = \Psi = \pm 180^\circ$ ).

#### 3.2. Metodología

Aunque, desde un punto de vista estrictamente formal, una proteína no es un objeto fractal, ya que no presenta una misma regularidad geométrica que sea invariante bajo cambio de escala, es posible estudiar su geometría, las irregularidades de su conformación, su grado de compactación, etc., mediante este tipo de geometría.

La geometría fractal ha sido utilizada como un método de cálculo para diferenciar las estructuras secundarias y terciarias de las proteínas, así como una herramienta que puede describir el grado de empaquetamiento adquirido por hélices  $\alpha$ , estructuras  $\beta$ , hélices  $\alpha_{3,10}$ , etc.

En este trabajo hemos cuantificado las conformaciones tridimensionales de la estructuras polipeptídicas  $\alpha$  y  $\beta$  generadas por ordenador mediante el software HyperChem [4].

Aminoácido N° de residuos	Dimensión Fractal ( $d_f$ ). Estructuras $\beta$ .									
	10	20	30	40	50	60	70	80	90	100
Serina (Ser)	1,0892	0,9213	1,0110	0,9458	0,9820	0,8818	0,9082	0,9223	0,9447	0,9506
Isoleucina (Ile)	1,1501	1,0723	1,2197	0,8881	0,9233	0,9401	0,9668	0,9834	0,9968	1,0084
Treonina (Thr)	1,1331	0,9551	1,1768	0,8625	0,9041	0,9122	0,9371	0,9553	0,9723	0,9784
Glutamato (Glu)	0,9534	1,0107	1,1061	1,0274	1,0723	0,9658	1,0116	0,0069	1,0274	1,0388
Lisina (Lys)	1,1649	1,069	1,1686	1,1673	1,0758	1,0983	1,1210	0,9988	1,0287	1,039
Aspartato (Asp)	1,1666	1,1434	1,2110	0,8904	0,9322	0,9449	0,9630	0,9816	0,9982	1,0029

**Tabla 2:** Dimensión fractal de diferentes modelos de estructuras  $\beta$ , obtenida según el número de residuos.

Por otro lado se han obtenido secuencias de proteínas así como sus coordenadas atómicas (definidas por difracción de rayos X y por Resonancia Magnética Nuclear) de la base de datos del Brookhaven National Laboratory, EUA (Protein Data Bank) [8].

Para calcular la  $d_f$  de la estructura de proteínas mediante el recuento por cajas hemos desarrollado un software que generaliza a 3 dimensiones ( $\mathbb{R}^3$ ) los conceptos explicados anteriormente. Concretamente el programa realiza los siguientes pasos:

1. Se genera un cubo de arista  $a$  en el que se inserta completamente toda la proteína o estructura secundaria objeto de estudio.
2. Se realizan divisiones sucesivas de la arista (dividiendo por la mitad), con lo que se van obteniendo 1, 8, 64, 512, ... cubos sucesivamente, y después de cada subdivisión se cuenta el número de cubos que contienen como mínimo un átomo de la proteína o estructura proteica.
3. Se representa el logaritmo del número de cubos contabilizados en el paso anterior en función del logaritmo del número de subdivisión. El resultado es una recta.
4. Se realiza un ajuste por mínimos cuadrados. La pendiente de la recta obtenida es la dimensión fractal de la estructura.

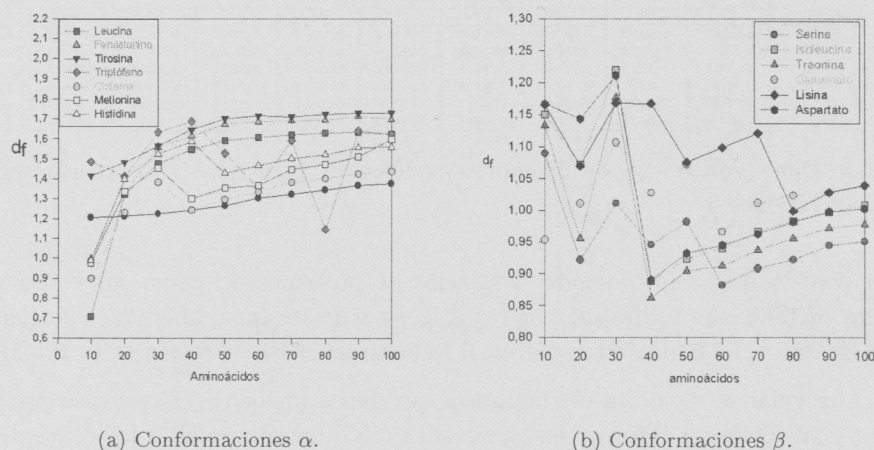
### 3.3. Resultados

#### 3.3.1. Cálculo de la $d_f$ de modelos polipeptídicos (hélices $\alpha$ y estructuras $\beta$ )

Para la primera parte del estudio se han generado hélices  $\alpha$  y estructuras  $\beta$  compuestas por unidades de 10 a 100 residuos y formadas por repetición de aminoácidos. Debido a la imposibilidad de generar verdaderas láminas  $\beta$  (diferentes cadenas polipeptídicas dispuestas una al lado de otra de forma paralela o antiparalela), se ha optado por generar lo que llamamos estructuras  $\beta$ , que corresponden a una única cadena de aminoácidos. Una vez realizados los modelos y obtenidas las posiciones atómicas se ha calculado la  $d_f$ . Los resultados obtenidos se presentan en las Tablas 1 y 2.

En la gráfica de la Figura 4 se ha representado la distribución de la dimensión fractal ( $d_f$ ) al variar el número de aminoácidos para las conformaciones  $\alpha$  y  $\beta$ .

Los resultados obtenidos ponen de manifiesto la relación existente entre el volumen de las cadenas laterales y la dimensión fractal en ambas conformaciones (hélices  $\alpha$  y estructuras  $\beta$ ).



**Figura 4:** Distribución de la dimensión fractal ( $d_f$ ) al variar el número de aminoácidos en conformaciones  $\alpha$  y  $\beta$ . En ambos casos se observa que al aumentar el número de aminoácidos, la dimensión fractal tiende a un valor constante, que llamamos dimensión fractal máxima ( $d_{f \text{ máx}}$ ).

Al aumentar el volumen de los aminoácidos, aumenta la  $d_f$ , comportamiento descrito por Daoud y Martin [1] para polímeros y que según los resultados obtenidos puede ser aplicado a las estructuras secundarias de proteínas.

En los dos modelos, a medida que aumenta el volumen de los aminoácidos se observa un empaquetamiento mayor, un mayor volumen y un aumento de la  $d_f$ .

La diferencias obtenidas entre las hélices  $\alpha$  y las estructuras  $\beta$  se deben a las conformaciones espaciales. Las hélices  $\alpha$  están más empaquetadas y las fuerzas de repulsión son mayores, mientras que en las estructuras  $\beta$  las repulsiones entre las cadenas son mucho menores y se encuentran menos empaquetadas. Por lo que la  $d_f$  de las hélices  $\alpha$  son mayores que la  $d_f$  de las estructuras  $\beta$ .

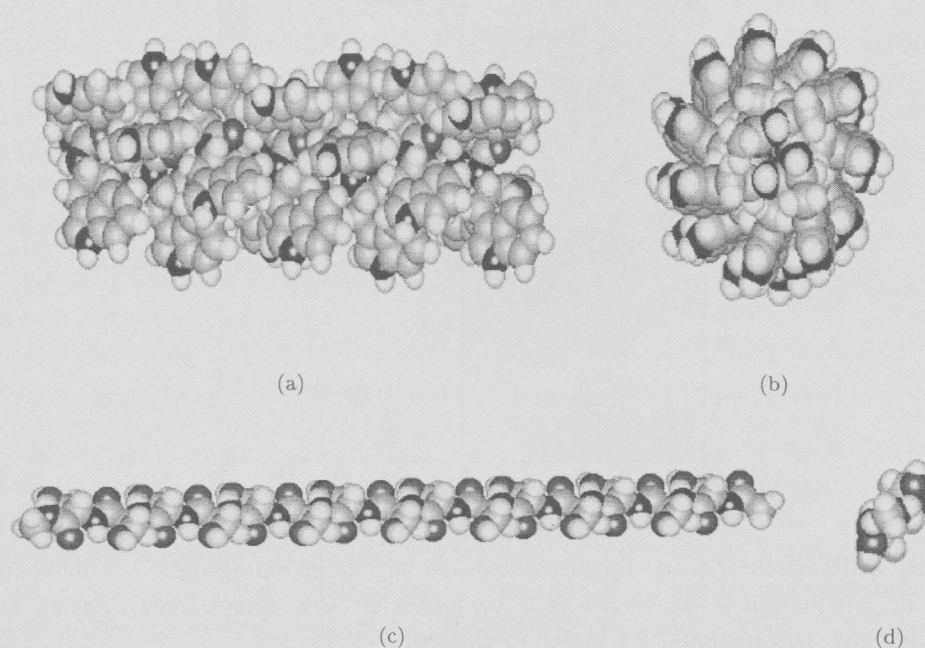
En la Figura 5 se representa un modelo de hélice  $\alpha$  formada por residuos de triptófano (Trp) y un modelo de una estructura  $\beta$  compuesta por residuos de serina generados por ordenador.

Para generar los modelos de hélice  $\alpha$  los ángulos  $\Phi(N-C_\alpha)$  y  $\Psi(C-C_\alpha)$  empleados en el estudio han sido:  $\Phi = \Psi = -135^\circ$ ,  $\Phi = \Psi = -90^\circ$ ,  $\Phi = \Psi = -45^\circ$ ,  $\Phi = \Psi = 0^\circ$ ,  $\Phi = \Psi = +135^\circ$ ,  $\Phi = \Psi = +90^\circ$  y  $\Phi = \Psi = +45^\circ$ . Los aminoácidos utilizados son: alanina (Ala), glicocola (Gly), leucina (Leu), fenilalanina (Phe), tirosina (Tyr), triptófano (Trp), cisteína (Cys), metionina (Met) e histidina (His).

La variación del valor de los ángulos  $\Phi$  y  $\Psi$  desde  $-135^\circ$  hasta  $+135^\circ$  genera hélices con distinto paso (o avance por residuo), creando hélices compactas en algunos casos, o hélices estiradas a modo de zig-zag en otros.

Las conformaciones de las cadenas polipeptídicas son estudiadas mediante la des-





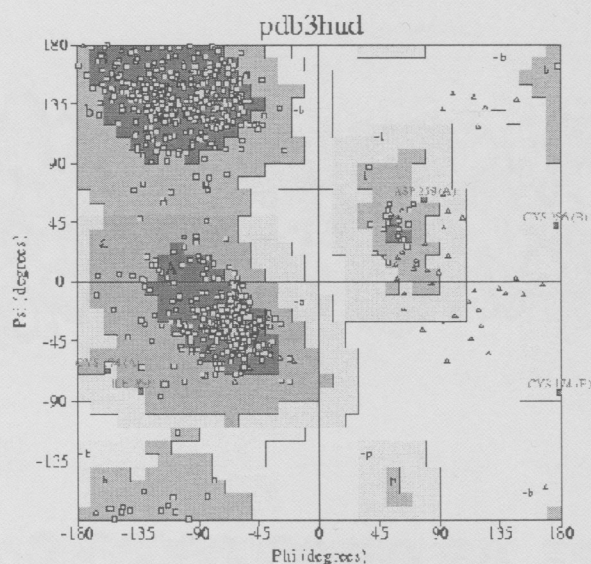
**Figura 5:** Hélice  $\alpha$  compuesta por residuos triptófano (modelos (a) y (b)), y estructura  $\beta$  formada por residuos serina (modelos (c) y (d)). En ambos casos se representa la vista lateral ((a) y (c)) y frontal ((b) y (d)).

cripción de cada uno de los residuos de aminoácidos en el gráfico bidimensional de Ramachandran. En el gráfico de la Figura 6 se sitúan los residuos entre los ángulos  $-180^\circ < \Phi < +180^\circ$  y  $-180^\circ < \Psi < +180^\circ$  y se muestran las conformaciones más normales observadas en las cadenas polipeptídicas.

En este trabajo el rango de ángulos utilizados ha sido ligeramente inferior a los rangos habituales en los diagramas de Ramachandran, pero se indica el intervalo de las áreas o zonas permitidas de mayor interés. Así, se incluyen en el estudio los ángulos de torsión  $\Phi$  y  $\Psi$  para las estructuras más importantes (hélices  $\alpha$  dextrógira, hélices  $\alpha$  levógira, estructuras  $\beta$  paralelas, estructuras  $\beta$  antiparalelas, hélices  $3_{10}$  y hélices del colágeno).

En los modelos estudiados se observa que aquellos que han sido generados con aminoácidos de bajo volumen (volumen estérico), como es el caso de la alanina y la glicocola, se obtienen  $d_{f\text{ máx}}$  para los ángulos  $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ . En cambio, para los modelos generados con aminoácidos con mayor volumen estérico, como el triptófano, cisteína, metionina, histidina y fenilalanina, se obtiene un  $d_{f\text{ máx}}$  para los ángulos  $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ .

¿Por qué los modelos generados por aminoácidos como Trp, Cys, Met, His y Phe no tienen  $d_{f\text{ máx}}$  a  $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ ?



**Figura 6:** Representación de Ramachandran del alcohol deshidrogenasa. Los diferentes tonos de gris indican la estabilidad de la conformación. Gris oscuro (muy estable), gris (estable), gris claro (permitido) y blanco (prohibido).

La respuesta puede encontrarse en el volumen ocupado por los grupos laterales o aminoácidos (volumen estérico) y en las interpretaciones dadas por Daoud y Martin [1] para el caso de los polímeros y que pueden ser aplicadas a las estructuras polipeptídicas estudiadas por nosotros.

En los modelos polipeptídicos generados con alanina y glicocola para los ángulos  $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ , se observa una  $d_f$  máx  $\approx 1,4$  ( $d_{f\text{ ala}} = 1,41$  y  $d_{f\text{ gly}} = 1,35$ ) que corresponde con el máximo empaquetamiento de la molécula.

La conformación más estable observada según los diagramas de Ramachandran corresponden a los modelos polipeptídicos generados con  $\Phi = \Psi = -45^\circ$  y que poseen mayor  $d_f$ , ya que sus residuos están situados en las zonas más favorables.

Para los polipéptidos generados con cadenas laterales de menor volumen estérico se observa cómo las conformaciones generadas con  $\Phi = \Psi = +45^\circ$  y  $\Phi = \Psi = -45^\circ$  se encuentran más empaquetadas y con menor distancia entre las cadenas laterales en cada una de las vueltas. A medida que el volumen estérico de las cadenas laterales aumenta, las repulsiones entre los grupos situados en vueltas contiguas aumenta y la conformación más estable es aquella que posee mayor paso o distancia entre residuos.

Los modelos creados con los aminoácidos fenilalanina, triptófano, metionina, histidina y cisteína obtienen  $d_{f\text{ máx}}$  para los ángulos de rotación  $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ .

Cuando el volumen de la cadena lateral es considerable provoca la mayor estabilidad del polipéptido a ángulos de rotación mayores que con la alanina y la glicocola.

La fenilalanina y el triptófano ( $V_{\text{Van der Waals}} = 135\text{\AA}^3$  y  $163\text{\AA}^3$ , respectivamente) poseen grupos laterales aromáticos. El triptófano tiene mayor volumen debido a su estructura bicíclica por lo que la  $d_{f\text{ máx}}$  del modelo generado con triptófano es mayor que la del modelo generado con fenilalanina. ( $d_{f\text{ máx}}(\text{Trp}) = 1,61$  y  $d_{f\text{ máx}}(\text{Phe}) = 1,56$ ). Las Figuras 7 y 8 representan las conformaciones tridimensionales de los dos modelos de polipéptidos generados con éstos aminoácidos. En ambos casos, la mayor compactación se produce a  $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ , pero la mayor estabilidad se produce para  $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ .

En los modelos de histidina, cisteína y metionina también se obtiene  $d_{f\text{ máx}}$  a los mismos ángulos de rotación debido al volumen de Van der Waals de las cadenas laterales. Para cada caso los valores son: metionina ( $V_{\text{Van der Waals}} = 124\text{\AA}^3$ ), histidina ( $V_{\text{Van der Waals}} = 118\text{\AA}^3$ ) y cisteína ( $V_{\text{Van der Waals}} = 86\text{\AA}^3$ ).

Al modelo generado con aminoácidos de metionina, aunque su cadena lateral sea lineal, el azufre contenido le confiere un volumen similar al de un grupo metileno, provocando que la cadena lateral sea parecida a una cadena ramificada, por lo que el volumen es elevado y la  $d_{f\text{ máx}}$  obtenida sea de 1,54. De la misma forma, el modelo de cisteína, que también contiene azufre en su grupo lateral, adopta las mismas propiedades estructurales que el modelo de metionina pero con una  $d_{f\text{ máx}} = 1,28$ .

El polipéptido formado por histidina se comporta de la misma forma que los cuatro modelos comentados anteriormente. Su anillo de imidazol le confiere un volumen de Van der Waals de  $118\text{\AA}^3$  y provoca, igual que en los casos anteriores, que la  $d_{f\text{ máx}}$  se obtenga para los ángulos  $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ . En el caso del modelo generado con histidina, por tener un volumen de Van der Waals inferior, la  $d_{f\text{ máx}}$  observada es ligeramente menor ( $d_{f\text{ máx}} = 1,59$ ).

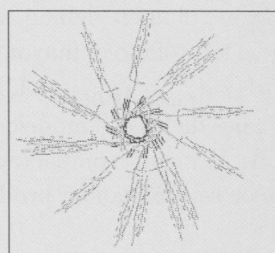
El último de los modelos polipeptídicos  $\alpha$  estudiados es el formado por aminoácidos de tirosina. Los resultados obtenidos con éste modelo no se ajustan a ninguno de los anteriormente citados. Adquiere una  $d_{f\text{ máx}}$  para ángulos de torsión  $\Phi = \Psi = -135^\circ$  y  $\Phi = \Psi = +135^\circ$ ; conformación en forma de zig-zag y casi plana.

Daoud y Martin [1] relacionan la dimension fractal con las fuerzas de atracción y repulsión (fuerzas de corto y largo alcance). En el caso de los aminoácidos con mayor volumen en las cadenas laterales (Leu, Phe, Tyr, Trp, Cys, Met y His), se crean repulsiones estéricas que hacen más estables las hélices con mayor paso (avance por residuo), de esta forma se justificarían las  $d_{f\text{ máx}}$  que hemos encontrado para estos ángulos.

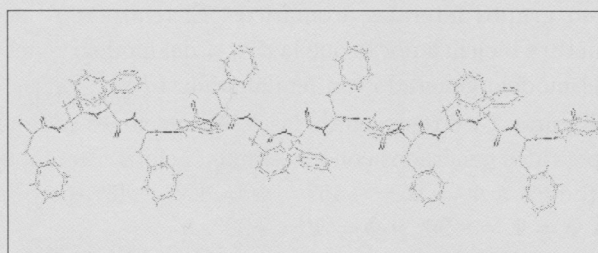
En los demás casos, ángulos ( $\Phi = \Psi = -135^\circ$  y  $\Phi = \Psi = +135^\circ$ ) y ( $\Phi = \Psi = 0^\circ$ ) se obtienen valores de  $d_f$  menores debido a la conformación que adopta una forma de hélice con un gran paso, para el primer caso, y una conformación casi plana y en forma de zig-zag para el segundo.

Por otro lado, las estructuras generadas que poseen mayor  $d_f$  tienen ángulos que se sitúan en zonas muy favorables sobre los diagramas bidimensionales de Ramachandran, por lo que la estabilidad que muestran es elevada.

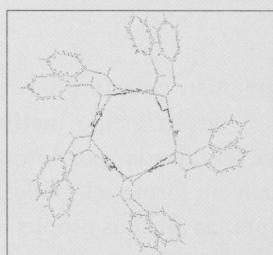




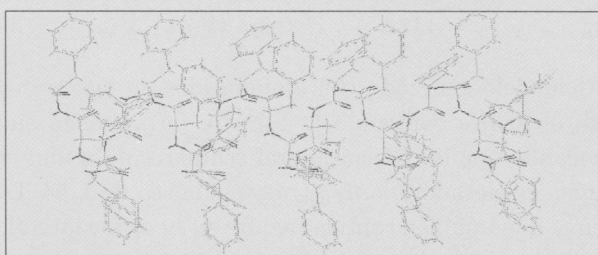
(a)  $\Phi = -135^\circ$



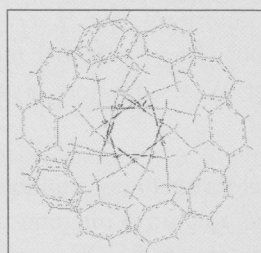
(b)  $\Psi = 135^\circ$



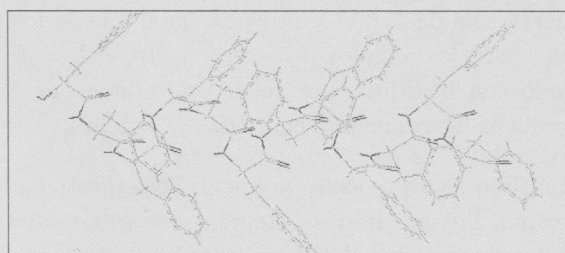
(c)  $\Phi = -90^\circ$



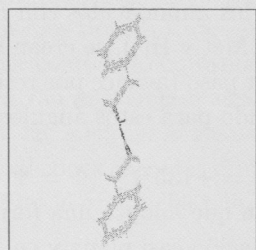
(d)  $\Psi = 90^\circ$



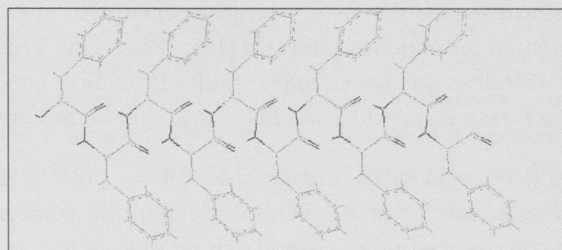
(e)  $\Phi = -45^\circ$



(f)  $\Psi = 45^\circ$

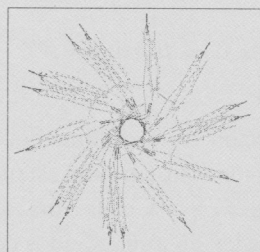


(g)  $\Phi = 0^\circ$

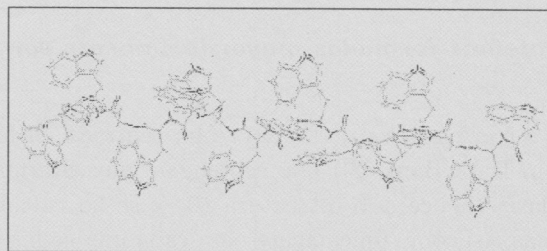


(h)  $\Psi = 0^\circ$

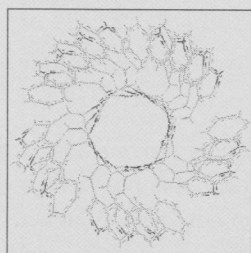
**Figura 7:** Conformación tridimensional de la polifenilalanina y ángulos de torsión.



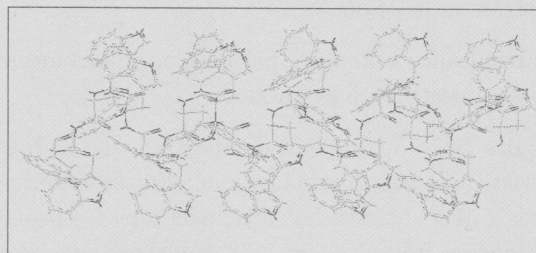
(a)  $\Phi = -135^\circ$



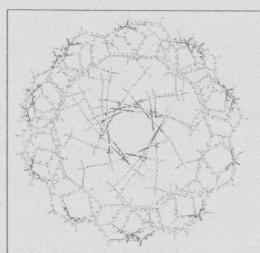
(b)  $\Psi = 135^\circ$



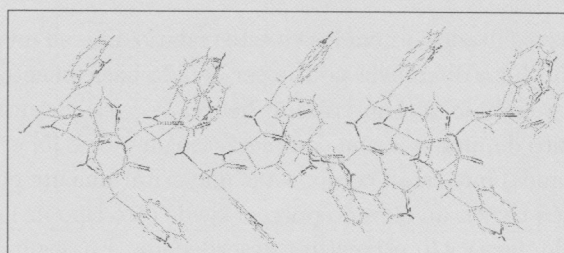
(c)  $\Phi = -90^\circ$



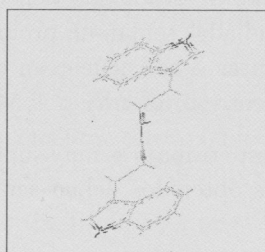
(d)  $\Psi = 90^\circ$



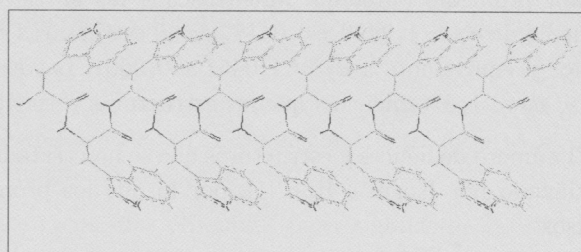
(e)  $\Phi = -45^\circ$



(f)  $\Psi = 45^\circ$



(g)  $\Phi = 0^\circ$



(h)  $\Psi = 0^\circ$

**Figura 8:** Conformación tridimensional del politriptófano y ángulos de torsión.

### 3.3.2. Cálculo de la $d_f$ de proteínas

#### *Proteínas formadas mayoritariamente por hélices $\alpha$ o láminas $\beta$*

La tercera parte del estudio se ha centrado en la selección de proteínas del PDB (Protein Data Bank [8]) que contengan mayoritariamente un sólo tipo de estructura secundaria (hélice  $\alpha$  o estructura  $\beta$ ), y se ha calculado la  $d_f$  para cada una de esas proteínas, así como únicamente para cada una de las secuencias de aminoácidos correspondientes a la conformación mayoritaria dentro de la proteína, descartando el resto de estructuras. Algunos de los resultados obtenidos para las más de 100 proteínas estudiadas se indican en las Tablas 3 y 4. Los resultados obtenidos para el caso de la hélice  $\alpha$  se muestran en la Tabla 3.

Los resultados obtenidos para las proteínas formadas mayoritariamente por estructuras  $\beta$  se muestran en la Tabla 4.

Isogai e Itoh [5] también demuestran que es posible asignar dimensiones fractales características a las proteínas que mayoritariamente están formadas por hélices  $\alpha$  o estructuras  $\beta$ . En sus estudios concluyen que las proteínas con conformaciones  $\alpha$  poseen una dimensión fractal  $d_f = 1,38$ , mientras que las proteínas con conformaciones mayoritariamente  $\beta$  presentan una dimensión fractal ligeramente menor ( $d_f = 1,29$ ). En nuestro caso los valores promedio encontrados son:  $d_f = 1,2490 \pm 0,0417$  para hélices  $\alpha$  y  $d_f = 1,0451 \pm 0,0601$  para estructuras  $\beta$ .

Los resultados obtenidos en este trabajo difieren un poco de los obtenidos por Isogai e Itoh [5]. Esta diferencia podría ser debida básicamente a dos motivos. En primer lugar, en este trabajo se han tratado hélices  $\alpha$  y estructuras  $\beta$  individualmente y no como conjunto de una proteína completa. Isogai e Itoh [5] estudian proteínas (mayor número de átomos) que están compuestas mayoritariamente por hélices o estructuras, pero que también tienen un elevado porcentaje de otro tipo de estructuras secundarias, como por ejemplo *turns* y/o estructura desordenada. En segundo lugar, el método utilizado en ambos trabajos es claramente diferente.

En la Figura 9 se muestra la  $d_f$  de cada una de las estructuras  $\alpha$  y  $\beta$  así como el valor promedio obtenido. Mientras que las 55 estructuras  $\alpha$  estudiadas tienen un promedio de  $d_f = 1,2490 \pm 0,0417$ , las 39 estructuras  $\beta$  estudiadas toman un promedio inferior y equivalente a  $d_f = 1,0451 \pm 0,0601$ . Las estructuras  $\alpha$  se encuentran entre  $0,91 < d_f < 1,67$ , mientras que las estructuras  $\beta$  se sitúan entre  $0,68 < d_f < 1,54$ .

El número de hélices y estructuras que se han tratado en este trabajo es muy superior al utilizado por Isogai e Itoh [5], por lo que los promedios obtenidos deben ser más precisos.

Isogai e Itoh [5] interpretan las diferencias de  $d_f$  entre las hélices y las estructuras debido a las fuerzas estéricas de repulsión y atracción entre los átomos vecinos, que es la misma conclusión a la que llegan Daoud y Martin [1]. Por su disposición en el espacio, las hélices sufren fuerzas de repulsión estérica entre sus cadenas laterales, mientras que en las estructuras con una disposición en zig-zag, las fuerzas estéricas son mucho menores.



PROTEÍNA	HÉLICE (secuencia)	PDB code	$d_f$	Coef. de correl.
DEDYROGENASE	HELICE 14 (313-334)	1GP1	1,27094	0,981
THERMOLISIN (HYDROLASE)	HELICE 4 (160-180)	8TLN	1,28156	0,991
OXIDOREDUCTASE	HELICE 1 (30-43)	1LLC	1,43765	0,9805
RIBUNUCLEASE (HYDROLASE)	HELICE 4 (13-29)	1SBS	1,21728	0,9645
PHOSPHOLIPASE (HYDROLASE)	HELICE 3 (39-58)	1BP2	1,27811	0,9851
FERROCYTOCHROME E.T	HELICE 5 (90-103)	1CYC	1,20749	0,9794
LEGHEMOGLOBIN	HELICE 7 (127-152)	1LH1	1,11243	0,9620
MYOGLOBIN	HELICE 8 (125-148)	1MBN	1,33171	0,9605
PHOSPHOLIPASE	HELICE 5 (89-108)	1P2P	1,22634	0,9882
RIBONUCLEASE	HELICE 1 (13-29)	1RNT	1,09796	0,9805
UTEROGLOBIN (OXIDIZED)	HELICE 1 (12-31)	1UTG	1,14564	0,8345
UTEROGLOBIN (OXIDIZED)	HELICE 2 (35-42)	1UTG	1,34563	0,9675
UTEROGLOBIN (OXIDIZED)	HELICE 3 (46-63)	1UTG	1,42648	0,9234
UTEROGLOBIN (OXIDIZED)	HELICE 4 (50-65)	1UTG	1,21911	0,9720
UTEROGLOBIN (OXIDIZED)	HELICE 5 (79-91)	1UTG	1,36343	0,9828
PREALBUMIN	HELICE 1 (75-83)	2PAB	1,36545	0,9763
DNA Binding Protein	HELICE 6 (94-105)	1WRP	1,22935	0,9685
LACTATE DEHYDROGENASE	HELICE 10 (309-330)	1LLC	1,09264	0,9604
LYSOZYME	HELICE 2 (25-35)	1LZ1	0,91175	0,9577
HYDROXYBENZOATE	HELICE 10 (299-310)	1PHH	0,97498	0,9827
SERINE PROTEINASE	HELICE 6 (220-238)	1SBC	1,09522	0,9839
TRIOSE PHOSPHATE ISOMERASE	HELICE 6-7 (130-154)	1TIM	1,21161	0,9753
TRYPTOPHAN SYNTHASE	HELICE 1 (1-12)	1WSY	0,96180	0,9657
PROTEINASE K	HELICE 5-6 (222-256)	2PRK	1,25791	0,9794
STAPHYLOCOCCAL NUCLEASE	HELICE 3 (122-306)	2SNS	1,18940	0,9586
CARBOXYPEPTIDASE ALPHA	HELICE 8 (285-306)	3CPA	1,43282	0,9756
GLUTATHIONE REDUCTASE	HELICE 3 (96-122)	3GRS	1,21582	0,9743
ACYL-COENZYME	HELICE 1 (3-15)	1ACP	1,49134	0,9345
ACYL-COENZYME	HELICE 2 (20-36)	1ACP	1,51668	0,9783
ACYL-COENZYME	HELICE 3 (51-62)	1ACP	1,44708	0,9234
ACYL-COENZYME	HELICE 4 (65-84)	1ACP	1,60125	0,9193
CHORISMATE MUTASE	HELICE 1 (6-41)	ECM	1,26262	0,9358
CHORISMATE MUTASE	HELICE 2 (49-66)	ECM	1,30265	0,9823
CHORISMATE MUTASE	HELICE 3 (70-94)	ECM	1,15012	0,9897
Poly(ADP-RIBOSE) TRANSFERASE	HELICE 1 (667-676)	PAX	1,04224	0,9012
Poly(ADP-RIBOSE) TRANSFERASE	HELICE 2 (679-688)	PAX	1,36906	0,9234
Poly(ADP-RIBOSE) TRANSFERASE	HELICE 3 (698-700)	PAX	1,12197	0,9649
LYSOZYME (HYDROLASE)	HELICE 1 (5-15)	1LYZ	1,15548	0,9134
LYSOZYME (HYDROLASE)	HELICE 2 (25-35)	1LYZ	1,21510	0,9536
LYSOZYME (HYDROLASE)	HELICE 3 (80-84)	1LYZ	0,91030	0,9985
LYSOZYME (HYDROLASE)	HELICE 4 (89-96)	1LYZ	1,25424	0,9782
PREALBUMIN	HELICE 1 (75-83)	2PAB	1,30917	0,9423
THERMOLYSIN (HYDROLASE)	HELICE 1 (65-87)	1TMN	0,99582	0,9872
THERMOLYSIN (HYDROLASE)	HELICE (137-150)	1TMN	1,34039	0,9372
THERMOLYSIN (HYDROLASE)	HELICE 4 (160-180)	1TMN	1,46947	0,9872
HYDROLASE SERINE PROTEINASE	HELICE 1 (234-245)	1EST	1,13202	0,9126
HYDROLASE SERINE PROTEINASE	HELICE 2 (164-170)	1EST	1,19867	0,9387
BACTERIORHODOPSIN	HELICE 1 (9-32)	1BAD	1,320744	0,9512
BACTERIORHODOPSIN	HELICE 2 (38-62)	1BAD	1,373844	0,9614
BACTERIORHODOPSIN	HELICE 3 (74-100)	1BAD	1,494062	0,9692
BACTERIORHODOPSIN	HELICE 4 (106-127)	1BAD	1,286396	0,9815
BACTERIORHODOPSIN	HELICE 5 (137-157)	1BAD	1,344876	0,9601
BACTERIORHODOPSIN	HELICE 6 (166-191)	1BAD	1,352454	0,9468
BACTERIORHODOPSIN	HELICE 7 (202-225)	1BAD	1,382335	0,9584
THERMOLYSIN	HELICE 4 (160-180)	8TLN	1,196348	0,9650

Tabla 3:  $d_f$  calculada para proteínas formadas mayoritariamente por hélices  $\alpha$ , y para diferentes hélices de estas proteínas.

### *Proteínas formadas por hélices $\alpha$ y láminas $\beta$*

De la misma forma que se ha procedido en el apartado anterior, y con el objetivo de complementar estos resultados, hemos estudiado la dimensión fractal de proteínas que contienen cantidades aproximadamente iguales de estructuras  $\alpha$  y  $\beta$ .

En la Tabla 5 se presentan los resultados obtenidos para algunas de las proteínas estudiadas, que en este caso contienen cantidades similares de estructuras  $\alpha$  y  $\beta$ .

Tal y como puede verse en la Figura 10, la  $d_f$  de las proteínas estudiadas se encuentra distribuida entre  $1 \leq d_f \leq 2$ . El valor promedio es  $d_f = 1,6186 \pm 0,0431$ . Los límites superiores e inferiores son  $d_f = 1,82$  y  $d_f = 1,42$ , respectivamente (para un intervalo de confianza del 95 %).

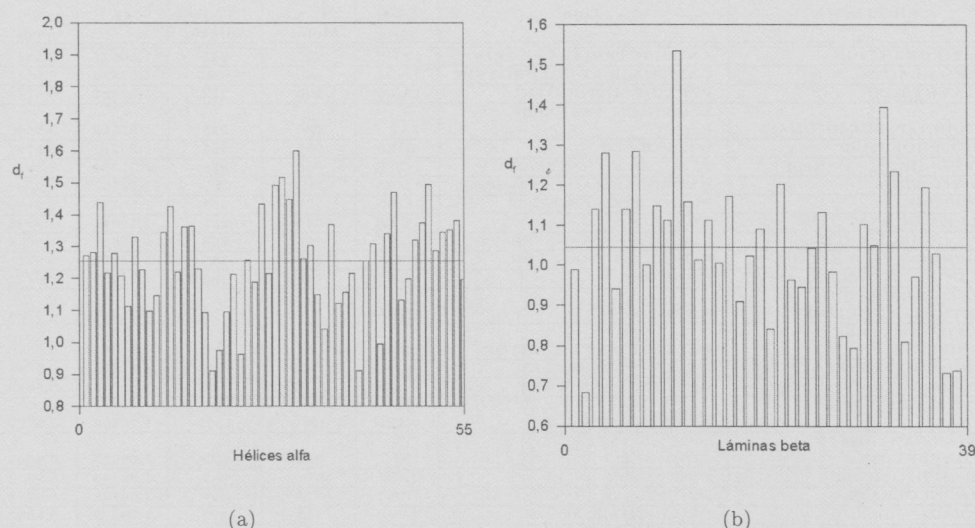
PROTEÍNA	LÁMINA (secuencia)	PDB code	$d_f$	Coef. de correl.
ACTINOXANTHIN ANTIBACTERIAL	LAMINA 1 (40-43)	1ACX	0,988237	0,9756
CHYMOSIN B (HYDROLASE)	LAMINA 2 (165-168)	1CMS	0,682698	0,9757
ALPHA COBRATION (TOXIN)	LAMINA 2 (20-25)	1CTX	1,139925	0,9183
FERRODOXIN ELECTRON TRANSPORT	LAMINA 2 (48-54)	1FDX	1,280913	0,9625
ALPHA AMYLASE INHIBITOR	LAMINA 4 (40-49)	1HOE	0,940583	0,9646
CONCANAVALIN A (LECTIN)	LAMINA 5 (108-116)	2CNA	1,140384	0,9728
CELLOBIOHYDROLASE (HYDROLASE)	LAMINA 2 (24-28)	1CBH	1,284775	0,9242
RIBONUCLEASE (HYDROLASE)	LAMINA 3 (94-104)	6RSA	1,001232	0,9467
RIBONUCLEASE (HYDROLASE)	LAMINA 2 (100-113)	7RSA	1,147957	0,9769
CARBONIC ANHYDRASE B	LAMINA 5 (140-151)	2CAB	1,112161	0,9921
CARBOXYPEPTIDASE B (HYDROLASE)	LAMINA 3 (24-28)	1CPB	1,533928	0,9733
TRYPSINOGEN HYDROLASE	LAMINA 7 (204-214)	1TGN	1,158067	0,9834
TONIN (HYDROLASE)	LAMINA 1 (41-48)	1TON	1,013085	0,9914
CARBONIC ANHYDRASE B	LAMINA 5 (140-151)	2CAB	1,112167	0,9921
PROTEINASE A (HYDROLASE)	LAMINA 2 (210-215)	2SGA	1,004065	0,9739
DIHYDROFOLATE REDUCTASE	LAMINA 7 (175-185)	8DFR	1,172594	0,9601
POLY(ADP-RIBOSE) POLYMERASE	LAMINA 1 (799-803)	1PAX	0,908503	0,9836
POLY(ADP-RIBOSE) POLYMERASE	LAMINA 2 (829-841)	1PAX	1,022171	0,9850
POLY(ADP-RIBOSE) POLYMERASE	LAMINA 3 (997-1009)	1PAX	1,091213	0,9793
PREALBUMIN (HUMAN)	LAMINA 1 (53-55)	2PAB	0,840912	0,9574
PREALBUMIN (HUMAN)	LAMINA 2 (105-112)	2PAB	1,201939	0,9589
PREALBUMIN (HUMAN)	LAMINA 3 (114-121)	2PAB	0,963035	0,9549
BACOCN (PAPIO CYNOCEPHALUS)	LAMINA 1 (40-43)	1ALC	0,945683	0,9672
BACOCN (PAPIO CYNOCEPHALUS)	LAMINA 2 (47-50)	1ALC	1,043423	0,9124
PHOSPHOLIPASE (HYDROLASE)	LAMINA 1 (74-78)	1BP2	1,132344	0,9234
PHOSPHOLIPASE (HYDROLASE)	LAMINA 2 (81-85)	1BP2	0,983453	0,9824
CRAMBIN (PLANT SEED PROTEIN)	LAMINA 1 (1-4)	1CRN	0,823434	0,9612
CRAMBIN (PLANT SEED PROTEIN)	LAMINA 2 (32-35)	1CRN	0,793547	0,9762
IMMUNOGLOBULIN (FGGI CLASS)	LAMINA 1 (239-243)	1FC1	1,102394	0,9776
IMMUNOGLOBULIN (FGGI CLASS)	LAMINA 2 (256-264)	1FC1	1,049543	0,9898
IMMUNOGLOBULIN (FGGI CLASS)	LAMINA 3 (299-309)	1FC1	1,394834	0,9723
IMMUNOGLOBULIN (FGGI CLASS)	LAMINA 4 (290-296)	1FC1	1,234374	0,9263
ACID PROTEINASE (HYDROLASE)	LAMINA 1 (1-7)	4APE	0,808234	0,9274
ACID PROTEINASE (HYDROLASE)	LAMINA 2 (13-22)	4APE	0,970012	0,9378
ACID PROTEINASE (HYDROLASE)	LAMINA 3 (24-33)	4APE	1,193845	0,9723
ACID PROTEINASE (HYDROLASE)	LAMINA 4 (117-124)	4APE	1,029345	0,9634
CYTOCHROME C3	LAMINA 1 (9-10)	2CY3	0,732165	0,9782
CYTOCHROME C3	LAMINA 2 (32-33)	2CY3	0,736267	0,9123

Tabla 4:  $d_f$  calculada para proteínas formadas mayoritariamente por estructuras  $\beta$ , y para diferentes secuencias de estructura  $\beta$  de estas proteínas.

Los resultados obtenidos en el presente trabajo son coincidentes con los estudios realizados por Li HQ y col. [3]. En su artículo: "Fractal analysis of protein chain conformation" se calcula la dimensión fractal de cinco proteínas: Lysozyme, Carboxypeptidase, Chymotrypsin ( $\alpha$ ), Myoglobin y Haemoglobin.

Estos resultados pueden ser interpretados de la misma forma que se interpretan las  $d_f$  de las estructuras secundarias (hélices y estructuras). Las proteínas que tienen una  $d_f$  inferior a 1,50 son aquellas que ocupan menos volumen y por lo tanto las fuerzas de atracción son mayores que las de repulsión. Las proteínas con una  $d_f$  superior a 1,50 ocupan mayor volumen y las fuerzas de repulsión son mayores que las de atracción. Y las proteínas con  $d_f \approx 1,5$  son las que poseen fuerzas de atracción y repulsión compensadas, teniendo en cuenta que las fuerzas son de largo y corto alcance.

Por otro lado, y según los resultados presentados en los dos apartados anteriores, la  $d_f$  obtenida en cada una de las proteínas está en relación directa al porcentaje de hélices  $\alpha$  y estructuras  $\beta$ . Las proteínas con  $d_f > 1,5$  son las que poseen mayor porcentaje de hélices, ya que por su distribución espacial aparecen más compactadas, ocupan mayor volumen. Las proteínas con  $d_f < 1,5$  tienen mayor proporción de estructuras  $\beta$  y las proteínas con  $d_f \geq 1,5$  tienen porcentajes proporcionales de hélices y estructuras.



**Figura 9:** Representación de la  $d_f$  de las proteínas estudiadas. La línea horizontal indica el promedio.

#### 4. Conclusiones

Desde un punto de vista estrictamente matemático las proteínas no son objetos fractales, pero es posible estudiar su estructura tridimensional utilizando este tipo de geometría.

Los resultados de este trabajo demuestran que es posible caracterizar y diferenciar las hélices  $\alpha$  y las estructuras  $\beta$  con un índice fractal diferente. Así mismo, se ha encontrado que las proteínas que están formadas mayoritariamente por estructuras  $\alpha$  tienen una dimensión fractal mayor que las que están formadas mayoritariamente por estructuras  $\beta$ .

En el estudio realizado mediante la generación por ordenador de modelos de polipéptidos  $\alpha$  y  $\beta$  con diferente número de aminoácidos se observa un aumento de la  $d_f$  a medida que aumenta el número de aminoácidos, hasta alcanzar un cierto valor constante. Las diferencias de  $d_f$  obtenida para cada uno de los modelos  $\alpha$  y  $\beta$  generados con distintos aminoácidos es debida al volumen estérico de la cadena lateral, propia de cada aminoácido. Se observa cómo a medida que el volumen estérico es mayor, la  $d_f$  también lo es.

Los modelos polipeptídicos  $\alpha$  y  $\beta$  generados obtienen  $d_f$  diferentes debido a los tipos de conformación tridimensional que adoptan cada uno. La  $d_f$  obtenida en los modelos polipeptídicos  $\alpha$  son mayores que la obtenida en los modelos  $\beta$ . Los modelos  $\alpha$  tienen estructuras más voluminosas, las hélices se encuentran más empaquetadas y las fuerzas de repulsión son mayores. Por el contrario, los modelos  $\beta$  adoptan conformaciones tipo zig-zag, en las que el volumen ocupado y el empaquetamiento es menor.

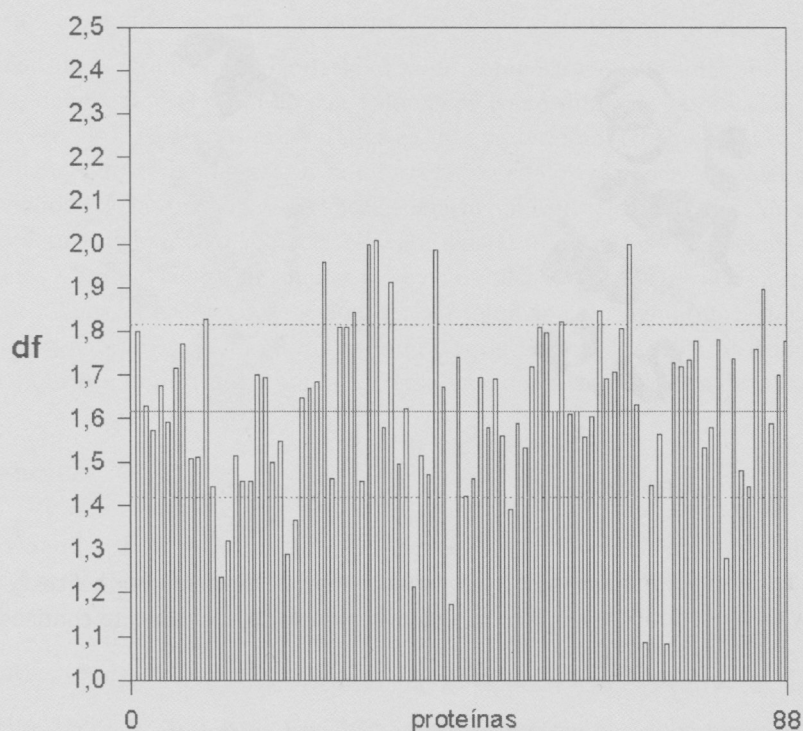


PROTEÍNA	Tipo	PDB code	Nº de átomos	Nº de enlaces	$d_f$	Coef. de correl.
ACTINOXANTHIN	ANTIBACTERIAL PROTEIN	1ACX	721	745	1,80027	0,9848
ALPHA-LACTALBUMIN	CALCIUM BINDING PROTEIN	1ACL	980	1005	1,62830	0,9639
ALAMETRICIN	PEPTIDE ANTIBIOTIC	1AMT	240	272	1,57491	0,9773
AZURIN (PSEUDOMONAS)	ELECTRON TRANSFER Copper	1AZU	930	945	1,67510	0,9846
PHOSPHOLIPASE	HYDROLASE	1BP2	957	944	1,59270	0,9767
CYTOCHROME C (RICE)	ELECTRON TRANSFER CY	1CCR	858	926	1,71594	0,9816
CHYMOSIN B	HYDROLASE(Acid Proteinase)	1CMS	2511	2574	1,77157	0,9794
CRAMBIN	PLANT SEED PROTEIN	1CRN	327	334	1,50747	0,9766
ALPHA COBRATOXIN	TOXIN	1CTX	541	556	1,51046	0,9843
CYTOCHROME C (Bonito)	ELECTRON TRANSFER CY	1CYC	795	868	1,82668	0,9801
FERRODOXIN (PEPTOCO)	ELECTRON TRANSFER	1FDX	373	385	1,44374	0,9711
GLUCAGON (HELICE ALFA)	HORMONE	1GCN	246	252	1,23480	0,9777
HISTOCOMPATIBILITY ANTIGEN A2 HUMAN	HORMONE	1HLA	367	0	1,31982	0,9734
ALPHA AMYLASE INHIBITOR	GLYCOSIDASE INHIBITOR	1HOE	558	570	1,51509	0,9685
LEGHEMOGLOBIN ACETAT	OXYGEN TRANSPORT	1LH1	1180	1218	1,45578	0,9645
LEGHEMOGLOBIN (AQUO)	OXYGEN TRANSPORT	1LH2	1180	1252	1,45623	0,9640
MYOGLOBIN (SPERM WHALE)	OXYGEN STORAGE	1MBN	1216	1292	1,70159	0,9720
AVIAN PANCREATIC POLYPEPTIDE	HORMONE	1PPT	602	620	1,69603	0,9569
PHOSPHOLIPASE	HYDROLASE	1P2P	971	933	1,49746	0,9819
RIBONUCLEASE	HYDROLASE	1RNT	777	824	1,54768	0,9778
MOL.IID	NEUROTOXIN	1SNB	482	492	1,28879	0,9032
UTEROGLOBIN (OXIDED)	STERIOD BINDING	1UTG	548	556	1,36642	0,9637
TRP REPRESOR TRIGONAL	DNA BINDING	1WRP	829	842	1,64671	0,9816
ALPHA BUGAROTOXIN	NEUROTOXIN	2ABX	1102	1133	1,66965	0,9903
BETA-LACTAMASE	HYDROLASE	2BLM	520	0	1,68521	0,9739
CYMOTRYPSIN INHIBITOR	PROTEINASE INHIBITOR CHYMOTRYPSIN	2CI2	521	528	1,95886	0,9467
CONCAVALIN A	LECTIN	2CNA	1251	1275	1,46292	0,9697
CYTOCHROME C3	ELECTRON TRANSPORT	2CY3	887	1096	1,80797	0,9723
INSULIN	HORMONE	2INS	778	800	1,81055	0,9618
INTERLEUKIN BETA	CYTOKINE	2ILB	1219	1242	1,84446	0,9731
LEGHEMOGLOBIN (AQUO)	OXYGEN TRANSPORT	2LH2	1180	1252	1,45578	0,9639
MYOHENERYTHRIN	OXYGEN TRANSPORT	2MHR	1999	2099	1,99876	0,9898
PREALBUMIN (HUMAN)	THYROXINE TRANSPORT	2PAB	1744	1795	2,00876	0,9747
PLASTOCYANIN	ELECTRON TRANSPORT	2PCY	738	753	1,58102	0,9670
MOL.IID	PHOTORECEPTOR	2PHY	101	1071	1,91148	0,9828
TOBACCO MOSAIC VIRUS	VIRUS	2TMV	1279	1311	1,49736	0,9744
CALMODULIN	CALCIUM BINDING PROTEIN	3CLN	1126	1138	1,62416	0,9681
CONCAVALIN A	LECTIN	3CNA	170	174	1,21241	0,9830
ELECTRON TRANSPORT	ELECTRON TRANSPORT	3FXN	1073	1122	1,51548	0,9624
VITAMIN D	CALCIUM BINDING PROTEIN	3ICB	600	613	1,47054	0,9537
INSULIN	HORMONE	3INS	1613	1654	1,98787	0,9738
TROPONIN C	CONTRACTILE SYSTEM	4TNC	1257	1265	1,67351	0,9769
ERABUTOXIN	TOXIN	5EBX	476	491	1,17234	0,9448
HIRUDIN	COAGULATION INHIBITOR	5HIR	671	673	1,74219	0,9590
BOVINE PANCREATIC TRYPSIN	PROTEINASE INHIBITOR	6PTI	458	490	1,42072	0,9509
CONCAVALIN A	LECTIN	2CNA	1251	1275	1,46292	0,9697

Tabla 5: Resultados obtenidos en proteínas con cantidades similares de estructuras  $\alpha$  y  $\beta$ .

En los modelos que han sido generados con aminoácidos que poseen grupos laterales poco voluminosos y con bajo volumen estérico, como es el caso de la glicocola y la alanina, la  $d_f$  máxima se ha obtenido para los ángulos ( $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ ). Sin embargo, en los modelos polipeptídicos  $\alpha$  generados con residuos de mayor volumen estérico, la  $d_f$  máxima obtenida coincide con los modelos generados con ángulos ( $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ ). Éste último caso es representativo de los modelos  $\alpha$  creados con fenilalanina, triptófano, metionina, histidina y cisteína.

La diferencia de la  $d_f$  máxima que se obtiene para los ángulos ( $\Phi = \Psi = -45^\circ$  y  $\Phi = \Psi = +45^\circ$ ) y ( $\Phi = \Psi = -90^\circ$  y  $\Phi = \Psi = +90^\circ$ ) puede interpretarse a partir de las observaciones realizadas por Daoud y Martin [1] para el caso de los polímeros. Estos autores identifican los valores obtenidos de  $d_f$  con las fuerzas de atracción y de

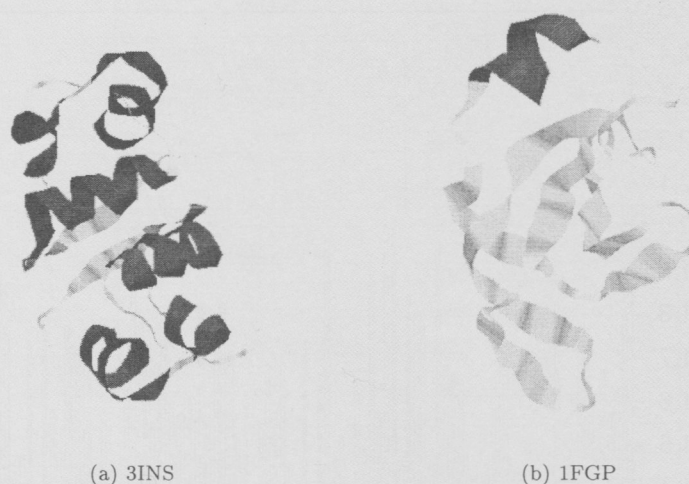


**Figura 10:** Dimensión fractal de las proteínas estudiadas. La línea continua indica el promedio de todos los valores. Las líneas discontinuas indican el promedio superior e inferior.

repulsión (tanto de corto como de largo alcance). Los modelos creados con aminoácidos cuyas cadenas laterales poseen gran volumen estérico (Leu, Phe, Tyr, Cys, Met y His), crean repulsiones estéricas que hacen más estable las hélices con mayor paso (avance por residuo), mientras que los modelos que poseen aminoácidos con cadenas laterales menos voluminosas, crean hélices más empaquetadas y con paso entre residuo menor.

En el mismo estudio se ha observado cómo todos los modelos  $\alpha$  medidos tienen  $d_f$  iguales para los ángulos de torsión ( $\Phi = \Psi$ ) de diferente signo. El resultado obtenido es lógico ya que la geometría fractal y la aplicación del método box-counting estudia su grado de irregularidad, considerando todas las escalas, y busca aspectos geométricos invariantes con el cambio de la misma. El hecho de obtener la misma  $d_f$  en modelos con ángulos de torsión  $\Phi = \Psi$  con signos opuestos demuestra que la medida de la  $d_f$  indica el grado de irregularidad y compactación y no da información sobre el ángulo de giro ( $d$  y  $l$ ) en la formación del polipéptido.

Los estudios realizados con estructuras secundarias (hélices  $\alpha$  y estructuras  $\beta$ ) seleccionadas entre las estructuras terciarias de las 88 proteínas resueltas mediante difracción de rayos X y Resonancia Magnética Nuclear (PDB), demuestran que cada una de estos tipos de estructura secundaria pueden caracterizarse con una dimensión fractal diferente.



**Figura 11:** El primer modelo (a) está compuesto mayoritariamente por hélices  $\alpha$  y alguna lámina  $\beta$  y su  $d_f > 1,5$  ( $d_f = 1,99$ ). El segundo modelo (b) únicamente contiene láminas  $\beta$  y su  $d_f < 1,5$  ( $d_f = 1,12$ ).

De las 55 hélices  $\alpha$  seleccionadas se ha obtenido un promedio de  $d_f = 1,2490 \pm 0,0417$ , mientras que de las 39 estructuras  $\beta$  el promedio obtenido es inferior,  $d_f = 1,0451 \pm 0,0601$ . Estas diferencias obtenidas para las estructuras secundarias de algunas proteínas se ajustan a los modelos  $\alpha$  y  $\beta$  generados de forma teórica, por lo que las diferencias entre la  $d_f$  de los dos tipos de estructura pueden interpretarse de la misma forma.

El estudio fractal realizado con las 88 proteínas completas seleccionadas del *Protein Data Bank* indica que la  $d_f$  se encuentra distribuida entre 1 y 2. El valor promedio obtenido es  $d_f = 1,6186 \pm 0,0431$ , con un margen de confianza del 95 %, resultados semejantes a los obtenidos por Li HQ y col. [3].

El mismo estudio muestra que es posible identificar proteínas que mayoritariamente están compuestas por estructuras  $\alpha$  o  $\beta$ . Aunque no se ha encontrado una relación directa entre el porcentaje de estructuras secundarias y la  $d_f$ , puede afirmarse que las proteínas con mayor proporción de estructuras  $\alpha$  tienen  $d_f$  mayores que las proteínas con estructuras  $\beta$ . Los promedios obtenidos en éste trabajo son:  $d_f(\alpha) = 1,6390 \pm 0,1350$  y  $d_f(\beta) = 1,5428 \pm 0,1727$ .

De los resultados estudiados se desprende que una primera interpretación del valor de la dimensión fractal debe tener relación con el grado de compactación de la estructura secundaria, y por tanto de la proteína. En este sentido hemos demostrado que las estructuras más compactas, más densas (hélices  $\alpha$ ) presentan valores de la dimensión fractal más elevados que las estructuras menos densas, menos compactadas, como las estructuras  $\beta$  (ver Figura 11). La misma interpretación debe hacerse en el caso de considerar proteínas enteras, y no sólo partes de ellas. Es posible, por tanto, asignar un



número que nos indica el grado de plegamiento de la cadena polipeptídica.

Puesto que el grado de irregularidad y de compactación de una proteína, de su superficie, debe estar relacionado con fenómenos como difusión, interacción con otras moléculas etc., los resultados presentados en este trabajo no deben ser considerados de forma aislada, sino que es necesaria su interrelación con otros aspectos relacionados con la funcionalidad de las proteínas. Se hace necesario por tanto, la realización de un estudio más completo y exhaustivo sobre la relación existente entre las dimensiones fractales encontradas y propiedades funcionales tales como plegamiento, interacción de ligandos con las proteínas, difusión, interacción de la proteínas con membranas, fármacos..., flexibilidad y vibración de las cadenas laterales de los aminoácidos, y cualquier otra propiedad que sea importante para la correcta funcionalidad biológica de las proteínas.

## Referencias

- [1] M. Daoud y J. E. Martin. Fractal properties of polymers. En D. Avnir, editor, *The fractal approach to heterogeneous chemistry. Surface, Colloids, Polymers*. John Wiley and Sons, New York, 1992.
- [2] Fractint (freeware fractal generator). <http://spanky.triumf.ca>.
- [3] Li HQ, Li Y., y Zhao HM. Fractal analysis of protein chain conformation. *Int. J. Biol Macromol*, 12(1):6-8, 1990.
- [4] Hyperchem™, 1993. Autodesk Inc.
- [5] Y. Isogai y T. Itoh. Fractal analysis of tertiary structure of protein molecule. *J. of the Physical Society of Japan*, 53(6):2162-2171, 1984.
- [6] B.B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman and Co., 1982.
- [7] B.B. Mandelbrot. *Los Objetos Fractales: forma, azar y dimensión*. Ed. Tusquets, 1993.
- [8] Protein data bank (pdb). <http://www.pdb.bnl.gov>. Brookhaven, EUA.
- [9] Pdbsum summaries and structural analyses of pdb data files. <http://www.biochem.ucl.ac.uk/bsm/pdbsum/>.
- [10] H.O. Peitgen, H. Jürgens, y D. Saupe. *Chaos and fractals: news frontiers of science*. Ed. Springer-Verlag, 1982.
- [11] H.J. Stapleton. Fractal form of proteins. *Physical Review Letters*, 45:1456, 1980.